Resolution Enhancement of Range Images
via Color-Image Segmentation

Arnav Bhavsar

PhD, Indian Institute of Technology Madras, India

Abstract

We propose a method for super-resolution of range image. Our approach leverages the interpretation of LR image as sparse samples on the HR grid. Based on this interpretation, we demonstrate that our recently reported approach which reconstructs dense range images from sparse range data by exploiting a registered colour image, can be applied for the task of resolution enhancement of range images. Our method only uses a single colour image in addition to the range observation in the super-resolution process. Using the proposed approach, we demonstrate super-resolution results for large factors (e.g. 4) with good localization accuracy.

1   Introduction

Within the last decade range cameras and scanners are established as an important dominant acquisition modality in computer vision and multimedia community. Applications where range cameras are successfully employed include digital heritage [1], industrial inspection, filming and 3DTV [2], gaming [3] etc. Range cameras operate on a variety of technologies involving laser scanning [4, 13], time-of-flight imaging [7, 8] and structured or coded lighting [12] which determine aspects such as accuracy, resolution, speed of acquisition and cost. For instance, there exists laser-based and structured-light-based scanners provide accurate and high resolution (HR) range scans, but the acquisition is typically slow and requires a good amount of manual effort. On the other hand, the low-cost laser scanners, time-of-flight scanners and light-coding range scanners (e.g. Kinect [3]) acquire range images much faster and with little manual intervention, but are limited in accuracy and resolution. Such a trade-off motivates the development of computational approaches to enhance the resolution and accuracy, so as to make range scanning more affordable, less time-consuming.

The low spatial resolution is a common issue with low-cost range scanning. Often, the acquired range image resolution is of the order of 100 to 200 pixels (in one direction) [7]. As compared to digital cameras (which typically span megapixels), these acquisition grids are too small to capture sufficient scene detail, edges etc. Hence, the resolution needs to enhanced by large factors (e.g. 4, 6 etc.). Naive approaches of enhancing the resolution by image interpolation methods results in heavy loss of accuracy. As a result, recent years has seen the rise of various sophisticated approaches [4, 5, 6] to enhance the resolution of range images which involve preserving edges as well as maintaining good accuracy.

A key to enhance the spatial resolution of range images by large factors, is to utilize a high-resolution registered colour image of the same scene, as done in these approaches. The colour image can be easily captured using any off-the-shelf digital camera. [1] The use of an HR colour image is motivated by an observation that commonly, the prominent depth discontinuities (such as those between different objects in a scene) coincide with colour discontinuities [14]. Thus, information about the presence/absence of colour discontinuities and local similarity of colour information, which is derived from the HR colour image helps in localizing range discontinuities and estimating dense range information on the HR grid. Such a cue from the colour image has also been used in many stereo disparity estimation approaches.

In this work, we notice that the problem of range super-resolution (SR) can, in fact, be interpreted as that of reconstructing a range image from sparsely captured data. In this context, we recently reported an approach from

---

[1] The registration can be achieved via one-time pre-calibration of the digital and range cameras or using any of the well established landmark-based registration approaches. Similar to the above referred approaches [4, 5], we too assume that the range and colour images are pre-registered.
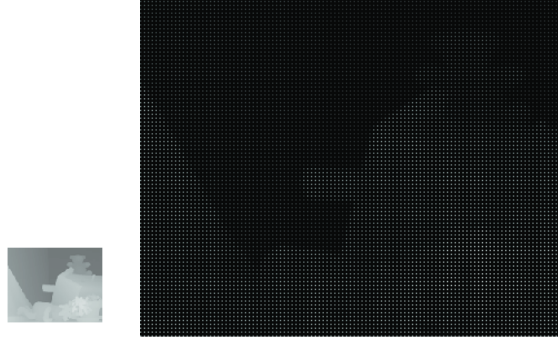
Figure 1: (a) LR range image and, (b) its corresponding sparse HR interpretation. The HR grid is $4 \times$ the LR grid in each direction.

reconstruction of dense range maps from sparsely captured range data [7], which uses a similar cue from the registered colour image via its segmentation. The advantages of this approach include a relatively simple local estimation approach, good accuracy, and computational efficiency. Here, we demonstrate that this approach would also hold for the task of range super-resolution.

The paper is organized as follows. In the next subsection we discuss some related work. Section 2 covers our methodology, which includes a brief description of the basic approach in [7]. We then provide some results in section 3, and conclude in section 4.

1.1 Related work

As mentioned earlier, the idea of using a colour image has been explored for range super-resolution in various ways. For example, the work in [16] interpolates the range image and, by exploiting the assumption that depth discontinuities coincide with color edges, improves estimation at discontinuities. Similar improvements are shown in MRF-based energy minimization approaches that also uses a HR optical image [15, 17]. However, these approaches are not known to perform for large super-resolution factors.

The authors in [4] propose an application of bilateral filtering which exploits constraints from the HR colour image. This work demonstrates ability to achieve good quality super-resolution by large factors. However, the bilateral filter, which is based on the HR colour image, has to be defined at each pixel. This makes the approach computationally very demanding. Our approach which works on image segments rather than pixels, and is based on computing local costs over segments, is a lot more efficient.

Another recent approach which also employs a colour image is reported in [5]. This method also uses the segmentation of the colour image. However, beyond this, it significantly differs from our method. The range labels are computed using a modified version of bilinear interpolation which is directed by the colour segmentation. On the other hand, ours approach computes range labels based on minimization of local costs with regularization constraints.

More recently, an example based range super-resolution approach is also reported [6], which also performs well at large resolution factors but requires a separate range dataset from which the training examples are derived. In contrast, our approach, like the above mentioned methods, only requires a registered colour image.

2 Resolution enhancement as reconstruction of sparse range data

As mentioned earlier, our approach is based upon the idea of treating the super-resolution problem as that of range reconstruction from sparse data [7]. This is inspired by the idea that LR range images can be essentially interpreted as down-sampled versions of their HR counterparts, where the down-sampling is obtained by selecting range pixels at regular intervals (depending on the SR factor). That is, the LR image is modeled as a decimation of the HR image. While in most image super-resolution approaches, the down-sampling is modeled by averaging of the HR pixels, the decimation model for down-sampling is not uncommon for range super-resolution [9, 8]. Thus, observing an LR image can be interpreted as observing only those samples of the HR image which were selected during the down-sampling process, with spaces between these samples having a zero value on the HR grid, thus resulting in a sparse HR image.

An example of the sparse interpretation of an LR range image is shown in Fig. 1, where for a resolution enhancement with a factor of 4, close to 93% of the data missing. Such large magnitudes of missing data are also considered in [7]. Hence, the SR task is essentially that of reconstructing such a sparse HR image.

2.1 Proposed method

As mentioned we begin by representing the low resolution range image on the high-resolution image grid by uniformly spacing the pixels in the low resolution range on the HR grid. The associated colour image is of the same size of this HR image grid.

The method starts with segmentation of the registered colour image. The colour segmentation is carried out using the well-known mean-shift algorithm (MSA) [10]. While a detailed description of the MSA is beyond the scope of this paper, it is worthwhile to state that the MSA uses two kernels defined on the colour dimension and the spatial dimension to determine the coarseness of the final colour segmentation. The greater these bandwidths, the coarser the segmentation and bigger the segments. These two parameters are important in the context of the approach in [7]. At the beginning, the method starts with small bandwidths and as a result achieves a largely over-segmented image, where each object is divided into multiple segments.

Following the initial colour segmentation, the approach computes plane-fit for each segment, based on the available range pixels in the segment. However, this is done only if the number of visible range pixels (n_a) that segment exceeds a threshold n_{pl} ($n_a \geq n_{pl}$). The plane-fitting is carried out using the RANSAC method [11]. Given the fitted-plane over a segment, a local cost is defined for assigning a range label z for each invisible pixel p in the segment as

$$C_p = |z - z_{pl}| + \lambda_p \sum_{q \in V_p} |z - z_q| \qquad (1)$$

Here, z_{pl} is the plane-fitted range at p and V_p is the set of *visible* second-order neighbours of p that belong to segment s. The first term computes distance from the plane-fitted range. The second term, weighted by λ_p, enforces similarity between neighbours. The label estimated for the pixel p is simply the one which minimizes C_p.

If for a segment $0 < n_a < n_{pl}$, then plane-fitting may not be robust enough. For such segments, a median range z_m over the n_a pixels is computed. In addition, the medians of the visible pixels over the adjacent segments are also computed. Based on this, a cost is defined as

$$C_s = |z - z_m| + w_a \sum_{z_{m_a} \in m_a} |z - z_{m_a}| \qquad (2)$$

where m_a is the set of the medians z_{m_a} of the visible pixels of $a \in A_s$. In equation 2, the second term enforces similarity over neighbouring segments. The weight w_a is defined for a pair of adjacent segments s and a as follows

$$w_a = \frac{\lambda_m}{(|\overline{r_s} - \overline{r_a}| + |\overline{g_s} - \overline{g_a}| + |\overline{b_s} - \overline{b_a}|)} \qquad (3)$$

Here, $\overline{r_i}$, $\overline{g_i}$, $\overline{b_i}$ are the mean RGB intensities for segment i. This contextual weighting strengthens the smoothness between similarly coloured segments but weakens it for segments with large colour differences. The label minimizing C_s is assigned to all the missing pixels in the segment s. Recall that this assignment is only for segments for which $0 < n_a < n_{pl}$. These are typically small segments for whom a constant surface assumption is quite valid.

As one would have observed, the above process only labels the pixels in the segments for which $n_a > 0$. However, there will be many segments with $n_a = 0$, when large number of pixels are missing. Just one pass of the above process, with a constant colour segmentation will leave many pixels unlabeled. To resolve this issue, the mean-shift segmentation, in each iteration, is performed with different (slightly larger) kernel bandwidths, so that as iterations progress, the segmentation produces somewhat larger segments. This process essentially subsumes many of the currently labeled and unlabeled segments into common larger segments. Thus, in subsequent iterations, many of the currently unlabeled segments are no longer isolated, thus making their pixels eligible for labeling. Since similar segments are more likely to be grouped, the demarcation between prominent range discontinuities that corresponds to largely different segments, is still maintained. Thus, the segment expansion process allows the cost computation modules to label all the pixels over iterations while maintaining prominent range discontinuities. A complete labeling is typically achieved in 4-5 iterations.
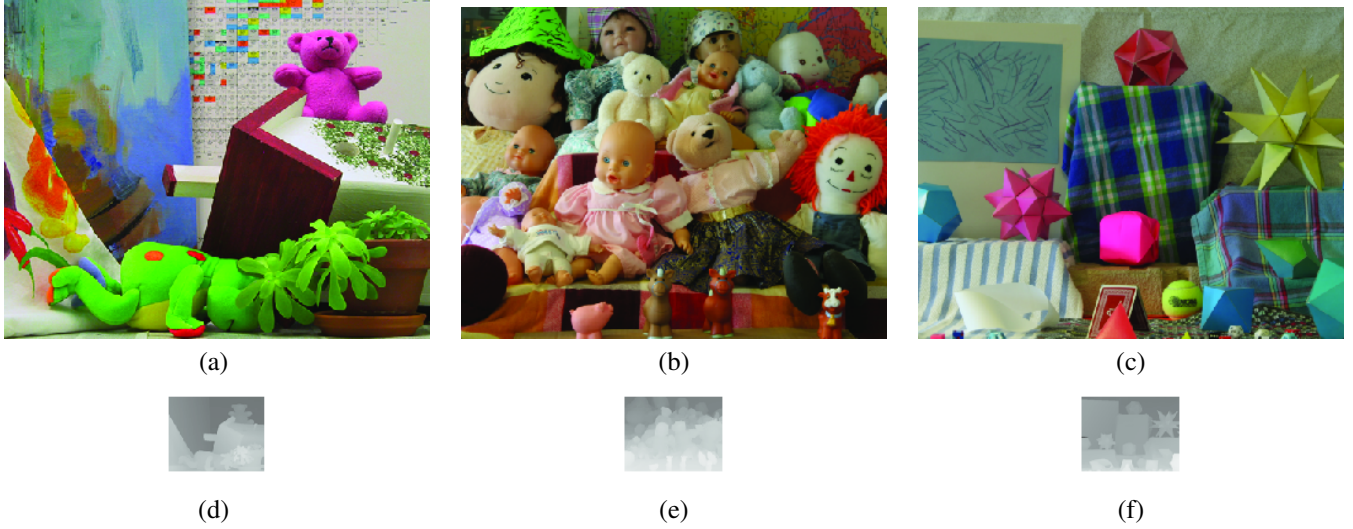
Figure 2: (a,b,c) Colour images from the Middlebury dataset used in our experiments. (d,e,f) Corresponding low resolution range images

3 Results

We now provide some results for our range super-resolution method described in the previous section. We conducted experiments with range images from the Middlebury dataset [12], which also has colour images associated with the range data. Some examples of the colour images corresponding to the scenes used in our experiments are shown in Figs. 2(a,b,c). The original resolution of the range images is of the order of 400×400 pixels. We down-sample these images and use the down-sampled versions as LR range images. We then compare our SR results with the original range images (or ground truth).

The results for the case of SR by a factor of 4 (Fig. 3). The low-resolution images are shown in Figs. 2(d,e,f). One can observe that it is difficult to distinctly perceive the image content in the LR images. Note that for a factor of 4, the bicubic-interpolated image is very blurred at the edges (Figs. 3(a,d,g)). Hence, the object shapes clearly lack proper localization. In comparison, our super-resolution result (Fig. 3(b,e,h)) shows clear improvements over the interpolated image. Such an improvement in the localization is the essence of range super-resolution. Moreover, in addition, one can also notice that high-level of fidelity when the super-resolved range images are compared with the ground-truth (Fig. 3(c,f,i)). As mentioned earlier, the factor of 4 case has a total of 93% missing data in its sparse interpretation. Inspite of this, our approach is able to maintain localization and fidelity in the eventual result.

Lastly, for completeness, we reassert that our super-resolution approach takes less than a couple of minutes for each of the above cases, on a Xeon 3.2 GHZ CPU with 12 GB RAM, with a Matlab implementation. Thus, we believe that our approach also performs well from the point of view computational efficiency. The efficiency of our method can be attributed to its local nature (i.e. there is no global energy minimization), as well as its segment-based (and not an explicit pixel-based) processing.

4 Conclusion

In this work, we investigated the problem of range super-resolution from the point of view of reconstructing dense range maps from sparse data. Based on the realization of LR images as sparse samples at the HR grid, we employed a recently proposed method for sparse range reconstruction, and demonstrated is applicability for the task of range super-resolution. Our approach shows promising results even for large resolution factors such as 4. In future, it would be interesting to gauge the performance of the approach and investigate further improvements under noise and non-exact registration between the range and colour image.
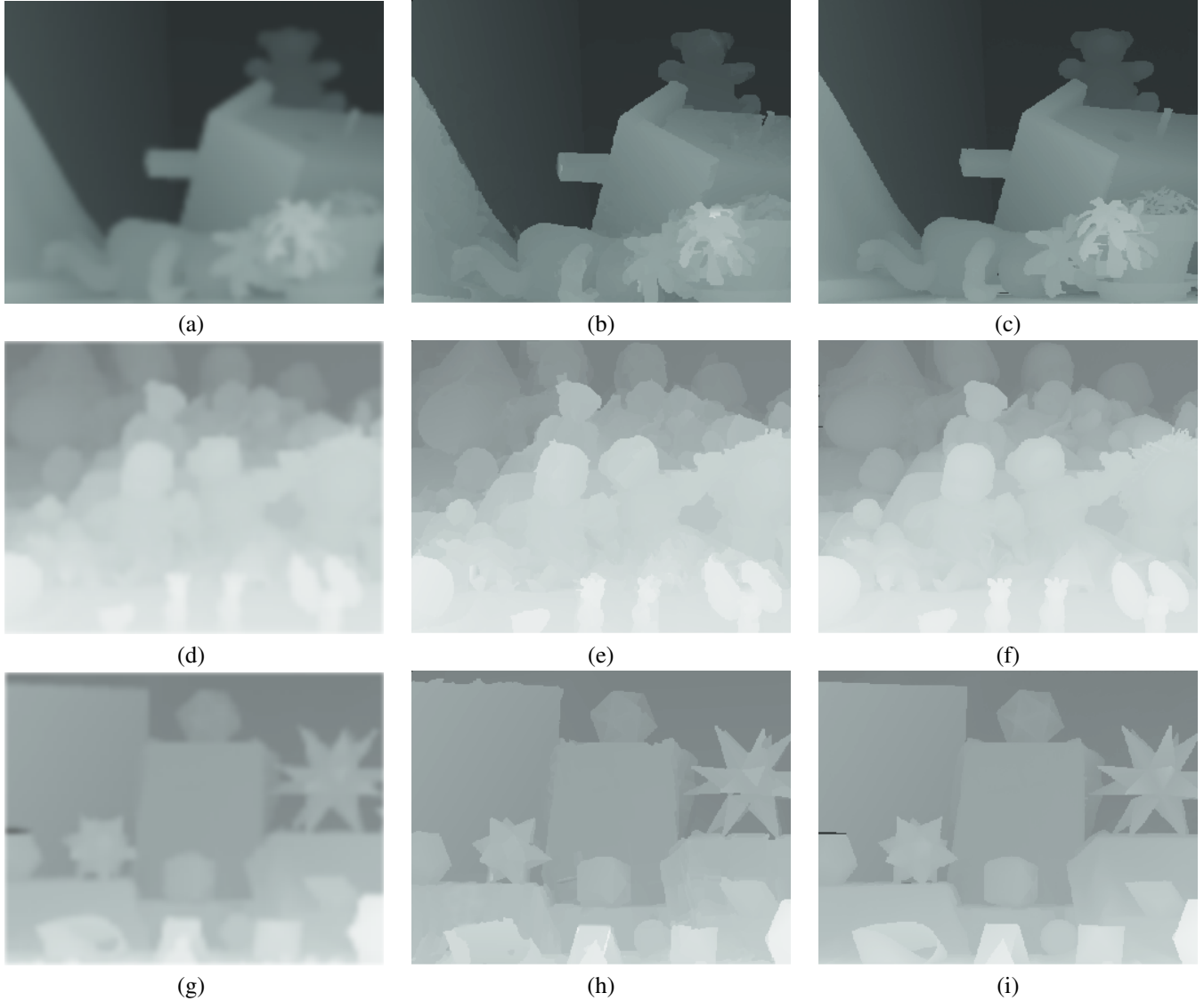
Figure 3: SR by a factor of 4: (a,d,g) Bicubic interpolation. (b,e,h) Super-resolved range image using our approach. (c,f,i) Ground-truth range.

References

[1] M. Levoy., *The digital Michelangelo project*, http://graphics.stanford.edu/projects/mich, (2009).

[2] B. Huhle, S. Fleck, S, A. Schilling., *Integrating 3D time-of-flight camera data and high resolution images for 3DTV applications*, 3DTV-Conference, pp. 1–4, (2008).

[3] J. Charles, and M. Everingham., *Learning shape models for monocular human pose estimation from the Microsoft Xbox Kinect*. IEEE Workshop on Consumer Depth Cameras for Computer Vision, 2011.

[4] Q. Yang, R. Yang, J. Davis, D. Nister., *Spatial-depth super resolution for range images*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008), pp. 1–8, 2008.

[5] V. Garro, P. Zanuttigh, G.M. Cortelazzo, *A New Super Resolution Technique for Range Data*, Proceedings of GTTI Meeting, 2009.

[6] O. Mac Aodha, N. Campbell, A. Nair and G.J. Brostow., *Patch Based Synthesis for Single Depth Image Super-Resolution*, European Conference on Computer Vision (ECCV 2012), 2012.

[7] A. Bhavsar and A. N. Rajagopalan, *Range map super-resolution-inpainting and reconstruction from sparse data*, Computer Vision and Image Understanding, vol. 116(4), pp. 572–591, 2012.

[8] S. Schuon, C. Theobalt, J. Davis, S. Thrun., *LidarBoost: Depth super-resolution for Tof 3D shape scanning*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009), pp. 343–350, 2009.

[9] S. Schuon, C. Theobalt, J. Davis, S. Thrun., *High-quality scanning using time-of-flight super-resolution*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008), pp. 1–7, 2008.

[10] C. Dorin, P. Meer., *Mean shift: A robust approach toward feature space analysis*, IEEE Trans. on Patt. Anal. and Mach. Intell., vol. 24(5), pp. 603–619, 1999.

[11] M. Fischler, R. Bolles., *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*, Communications of the ACM, vol. 24(6), pp. 381–395, (1981).

[12] D. Scharstein, R. Szeliski., *High-accuracy stereo depth maps using structured light*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2003), vol. 1, pp. 195–202, (2003).

[13] Y. Kil, B. Mederos, N. Amenta., *Laser scanner super-resolution*, Eurographics Symposium on Point-Based Graphics, pp. 9–16, 2006.

[14] Q. Yang, L. Wang, R. Yang, H. Stewenius, D. Nister., *Stereo matching with color-weighted correlation, heiarchical belief propagation, and occlusion handling*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 31(3), pp. 492–504, 2009.

[15] B. Huhle, S. Fleck, S, A. Schilling., *Integrating 3D time-of-flight camera data and high resolution images for 3DTV applications*, 3DTV-Conference, pp. 1–4, 2008.

[16] H. Andreasson, R. Triebel, A. Lilienthal., *Vision based interpolation of 3D laser scans*, Proceedings of the International Conference on Autonomous Robots and Agents, 2006.

[17] J. Diebel, S. Thrun., *An application of Markov random fields to range sensing*, Conference on Neural Information Processing Systems (NIPS 2005), 2005.